Lesson 6

Analysis and Interpretation of Surveillance Data

Instructor's Guide Form

Lesson Title: Analysis and interpretation of surveillance data

Lesson Goal: For each student to be able to describe the approaches to analysis and interpretation of surveillance data.

Lesson Objectives: By the end of this lesson, the learner will be able to:

1) describe the methodological approaches to surveillance analysis
2) describe the practical approaches to surveillance analysis
3) present surveillance data by time, place, and person
4) describe the concept of rates and standardization of rates
5) describe the approaches to exploratory data analysis
6) demonstrate the uses of graphics and maps
7) describe the systematic interpretation of surveillance data

Equipment and materials needed:

- Overhead projector
- Transparencies #6.1 - #6.57

Lesson 6

Analysis and Interpretation of Surveillance Data

Instructor's Guide Form
(continued)

- Optional exercise - Interpretation of surveillance data: "Epidemiologic Patterns of Four Acute Communicable Diseases" (Time required 2-3 hours) - *See end of lesson*

Time required: 90 minutes

Synopsis of lesson: This lesson provides the methodology and practical approaches to the analysis and the interpretation of surveillance data.

Adult Education Application: Adults want to practice the skills taught to them and opportunities to adapt these skills for use in their job settings. The ability to format surveillance data into graphs, charts, and figures is an example of one of these skills. Rather than just reviewing the uses of various graphic presentations to the learners, you could set up a typical public health situation that requires them to format the same surveillance data into different presentations. Once they have arranged the data into graphs, charts, and figures, you can lead a discussion that highlights the benefits and limitations of each format. You could add a further discussion on how each of these methods might be used with different audiences, such as in a scientific meeting, political setting, or an academic classroom.

Lesson 6

Analysis and Interpretation of Surveillance Data

Topical Outline

I. Introduction
- A. Importance of analysis and interpretation

II. Approaches to surveillance analysis
- A. Practical approach
- B. Methodological considerations
- C. Time, place, person
- D. Interactions among time, place, person

III. Rate and rate standardization
- A. Rate
- B. Use of rates in epidemiology
- C. Crude, specific, and standardized rates
- D. Analysis of rates

IV. Exploratory data analysis (EDA)
- A. EDA
- B. Techniques for data displays
- C. Types of data summaries
- D. Transformations
- E. Smoothing

V. Data graphics
- A. Purpose of graphics
- B. Tables
- C. Graphs
- D. Maps

Lesson 6

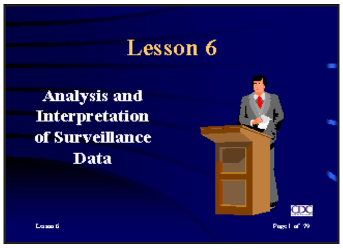
Analysis and Interpretation of Surveillance Data

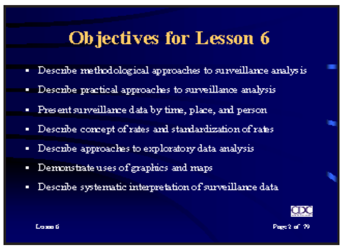
Topical Outline
(continued)

VI. Interpretation of surveillance data
- A. Factors in interpreting data
- B. Limitations in data
- C. Approach to interpretation

VII. Interpretative uses for surveillance data
- A. Identify epidemics
- B. Identify new syndromes
- C. Monitoring trends
- D. Evaluating public policy
- E. Projecting future needs

Click on thumbnail to see full image





Lesson 6
Analysis and Interpretation of Surveillance Data

Content Outline

Lesson Objectives

- **Describe methodological approaches to surveillance analysis**
- **Describe practical approaches to surveillance analysis**
- **Present surveillance data by time, place, and person**
- **Describe concepts of rates and standardization of rates**
- **Describe approaches to exploratory data analysis**
- **Demonstrate uses of graphics and maps**
- **Describe systematic interpretation of surveillance data**

I. Introduction

A. Importance of analysis and interpretation

1. whether surveillance is used to characterize trends in disease or injury, detect epidemics, suggest hypotheses, evaluate prevention programs, or project future public health needs, data must be analyzed carefully and interpreted prudently

2. ability to effectively analyze, interpret, and present surveillance data is an important skill for the public health worker

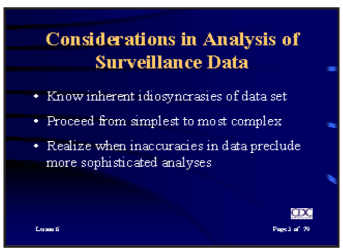
Return to Lesson 6

II. Approaches to surveillance analysis

A. Practical approach

1. surveillance data can be used for: monitoring trends, portraying the natural history, identifying epidemics and new syndromes, monitoring changes in infectious agents, identifying areas for research, evaluating hypotheses, planning, evaluating public policy, and evaluating public health interventions
2. surveillance analysis represents an inductive reasoning process in which the assembly of individual units eventually produces a more general picture of health-related problems in a population

3. considerations in analysis of surveillance data



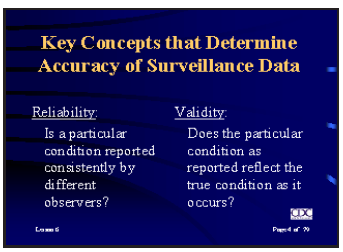
   a. knowledge of the strengths and weaknesses of data collection methods and reporting process can provide a "real world" sense of the trends that emerge
   b. proceed from simplest to the most complex
      1) examine each condition separately, by numbers and crude trends
      2) absolute numbers provide a crude evaluation of the condition's occurrence
      3) determination of rates is a more accurate evaluation of the condition's occurrence and accounts for denominator changes (see IIIA below)
      4) look at variables separately prior to examining relationships among variables

Return to Lesson 6

c. realize when inaccuracies in the data preclude more sophisticated analyses
    1) differential reporting by different regions or sites render resulting data set liable to misinterpretation
    2) erratically collected or incomplete data cannot be corrected by complex analytic techniques
    3) nonrepresentative data precludes accurate analysis

4. surveillance data cannot be used for formal hypothesis testing

B. Methodological Considerations



1. key concepts that determine accuracy of surveillance data
    a. reliability
    b. validity

2. reliability
    a. is a particular condition reported consistently by different observers?
    b. is the information reproducible?

3. validity
    a. does the particular condition, as reported, reflect the true condition as it occurs?
    b. more difficult to assess than reliability

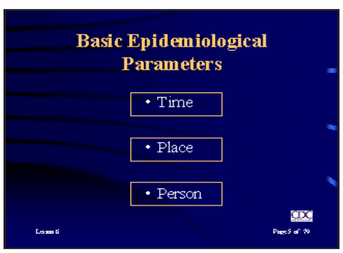
4. application of standard statistical techniques to analysis of surveillance data
    a. dictated by limitations of data
    b. if information is viewed as samples over time, apparent clusters of health events can be evaluated for their statistical “significance”

Return to Lesson 6

1) apply 95% confidence limits or other standard statistical tests to samples over time
2) can allow a determination of whether any differences are unlikely to have occurred by chance alone

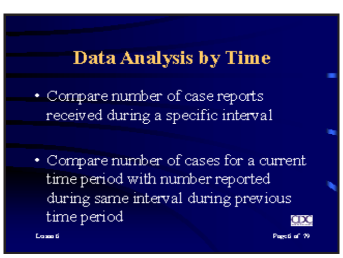
5. ecological fallacy
   a. may occur when interpreting observations about groups
   b. problem of ecological fallacy
      1) aggregation bias may arise from loss of information when individuals are grouped
      2) specification bias may arise from the definition of the group itself
   c. can reduce chances of ecological fallacy by analyzing subsets of surveillance data to reveal trends in individual characteristics of survey data

C. Time, place, person



1. surveillance data allows the description of health problems in terms of basic epidemiologic parameters (time, place, person)

2. surveillance data permits comparisons among these different parameters (time, place, person)
3. use of these data permits
   a. long term trends to be monitored
   b. seasonal patterns to be assessed
   c. epidemics to be detected
   d. future occurrence of disease or injury to be projected



4. time
   a. analysis of surveillance data by time can reveal trends in disease and injury

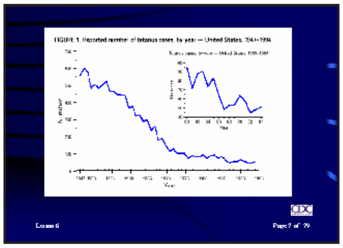
Return to Lesson 6

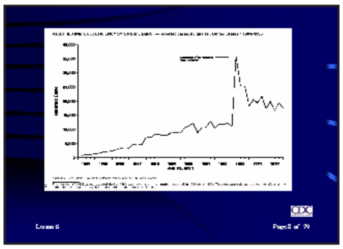
1) methods of analysis of surveillance data include:
    a) comparison of number of case reports received during a particular interval (days, weeks, months)
    b) can be organized into a table or graph
    c) comparison of number of cases for a current time period with number reported during the same interval for the past several years
    d) using denominators to calculate rates becomes especially important if changes occur in population
    e) analysis by date of onset rather than date of report more clearly defines condition

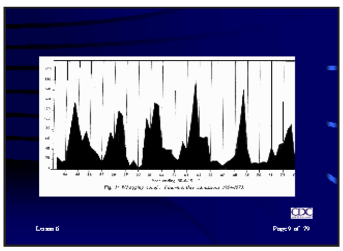
b. the analysis should explain:
    1) whether trends are stable, gradually rise and fall, or show occurrence of an abrupt increase over that which was expected
    2) long term secular trends

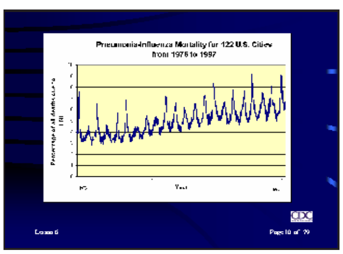
c. limitations to data analysis include:
    1) a measurable delay occurring between the exposure and reporting of the problem
    2) identify which interval concerning the case will be used
        a) between exposure and expression of symptoms
        b) between onset of symptoms and diagnosis of problem
        c) between diagnosis and reporting of illness to public health authorities so it can be included in surveillance data set

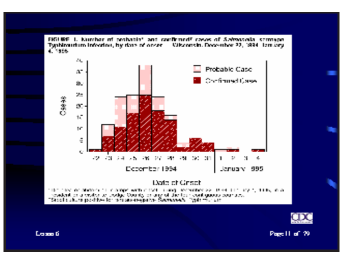
Return to Lesson 6

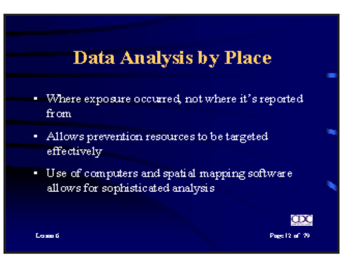












3) choosing the appropriate interval for analysis must involve a consideration of the health condition being assessed

d. examples of time analyses:

1) secular trend (tetanus for example).

a) events that influence secular trends can be portrayed on the graph

1) changes in case definition used for surveillance, new diagnostic criteria, changes in reporting requirements or practices, publicity about a particular condition, or increased intensity of case detection

2) example: case definition, AIDS

3) example: publicity, Lyme disease

4) example: case detection - measles

b) changes in surveillance system can influence long-term trends

2) periodicity such as seen in herd immunity, example - Pertussis

3) seasonal trend, example-Influenza

4) epidemic, Samonella example

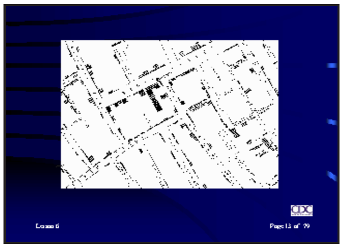
5. place

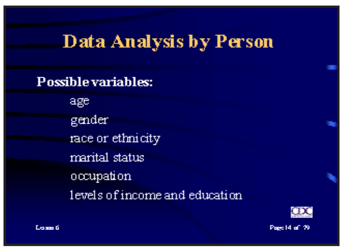
a. place is where the exposure occurred, which may not be from the same place reported

b. size of unit for geographic analysis is determined by type of condition involved

c. may portray data by table or spot map

Return to Lesson 6





d. computers and spatial mapping software allow for more sophisticated analysis of surveillance data by place
e. must locate the geographic area with the highest rates to facilitate efforts to identify causes and allow appropriate interventions to be applied
f. identification of focal areas allows prevention resources to be targeted effectively
g. John Snow's removing Broad Street pump handle is classic example of intervention by location

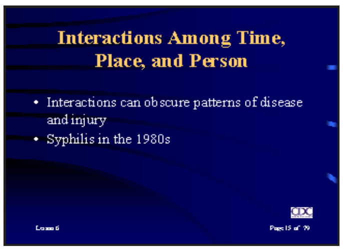
6. person
   a. types of possible demographic variables includes
      1) age
      2) gender
      3) race or ethnicity
      4) marital status
      5) occupation
      6) levels of income and education
   b. analysis of trends in disease or injury by age depends on specific health condition of interest
      1) for childhood diseases, relatively narrow age categories can identify age group associated with peak incidence of a particular health condition (for example, measles)
      2) for conditions which affect older age groups, broader 10-year age intervals are frequently used
      3) best guide to deciding which age categories to use is the typical age distribution associated with the health condition
   c. analysis of behavioral characteristics of populations

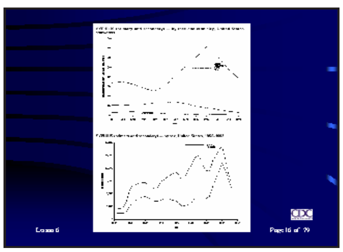
Return to Lesson 6

1) depend on self-reported behavior
2) may be based on repeated surveys of representative groups, trend in markers for specific types of behavior, or active surveillance of a particular behavioral characteristic or indicator in a defined group

d. characteristics of persons in surveillance system should be related to denominators, if possible
e. example: age data reveals tetanus is an adult problem in U.S.A.

D. Interactions among time, place, person

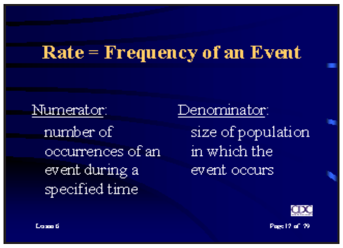




1. interactions can obscure important patterns of disease and injury in specific populations, for example:
   a. in U.S. in 1980s, overall number of syphilis cases fell during first two-thirds of decade but began to rise in 1987
   b. when analyzed by gender, decline in syphilis occurred primarily among men
   c. when stratified by gender and race, decrease in numbers of cases of syphilis was seen only among white men, presumably among men who had sex with other men and who changed sexual practices in response to HIV prevention practices
   d. increase in syphilis occurred among black men and women and was linked to unsafe sexual behavior associated with use of crack cocaine

2. without stratified analysis by person, offsetting trends in mid-1980s of declines among white males might have delayed recognition by public health officials of the syphilis epidemic among minorities

Return to Lesson 6



III. Rate and rate standardization

A. Rate

1. rate measures frequency of an event

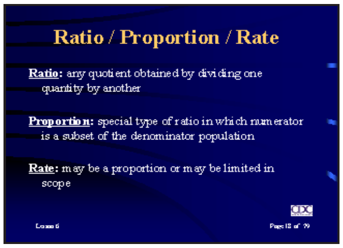
2. comprises a numerator and a denominator
   a. numerator denotes number of occurrences of an event during a specified time
   b. denominator denotes size of population in which the event occurred
   c. denominator determined by numerator data (e.g. if analyzing annual occurrence data, denominator generally is the estimated mid-year (July) population)

3. specification of time period under consideration is crucial

4. multiplier (a power of 10 used to convert awkward fractions to more workable numbers 100 - 1,000,000)

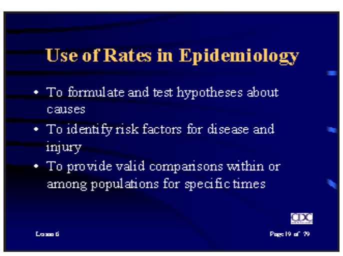
5. use of raw numbers
   a. raw numbers provide a valuable immediate analysis of data
   b. raw numbers quantify occurrences of an event during a specified time without regard to population size and dynamics or other demographic characteristics such as distribution by race and gender
   c. rates enable one to make more appropriate, informative comparisons of occurrences in a population over time, among different sub-populations, or among different populations at same or different times, since size of population and period of time specified are accounted for in calculation of rates

Return to Lesson 6



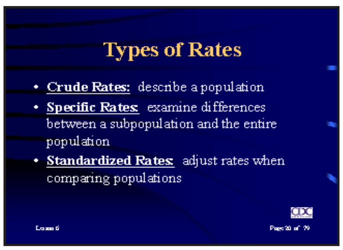
6. differences between ratio, proportion, rate
    a. ratio
        1) any quotient obtained by dividing one quantity by another
        2) numerator and denominator are generally distinct quantities, neither of which is a subset of the other
        3) no restrictions exist on the value or dimension of a ratio
        4) allows the comparison of one quantity with another quantity
    b. proportion
        1) special type of ratio
        2) numerator is a subset of the denominator population
        3) requires resulting quotient to be dimensionless, positive, and less than one, or less than 100 if expressed as a percentage
    c. rate
        1) measures risk
        2) in epidemiology, a rate may be a proportion (such as a prevalence rate) or may be limited in scope by further restrictions such as representing the number of occurrences of a health event in a specified time and population per unit time

B. Use of rates in epidemiology



1. purposes of calculation and analysis of rates
    a. to measure magnitude of problem
    b. to formulate and test hypotheses about causes
    c. to identify risk factors for disease and injury
    d. to provide valid comparisons within or among populations for specific times

Return to Lesson 6



2. must have reliable numerator and denominator data to determine rates
    a. if data to be analyzed have been collected from public health surveillance systems, denominator data can be difficult to obtain
    b. small numerators and / or denominators can cause rates to fluctuate

C. Crude, specific, and standardized rates

1. crude and specific rates
    a. crude rates
        1) rates which describe a complete population
        2) computation of crude rates is performed as initial step in analysis
        3) crude rates are important in obtaining information about and contrasting entire populations
        4) example: crude mortality rate: number of deaths from all causes divided by total population times 1000
    b. specific rates
        1) used to examine differences between a subpopulation and the entire population
        2) population is partitioned into relevant “specific” subpopulations and a “specific” rate is calculated for each subset
        3) example: disease specific mortality rate: deaths due to a specific disease divided by the total population times 100,000

Return to Lesson 6

c. why is there variation of rates among population groups?
    1) natural history of health problem
    2) differential distribution of susceptibility or causes
    3) genetic differences among subpopulations

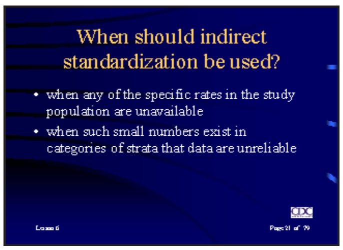
2. standardized rates
    a. crude rates are appropriate only if populations are similar with respect to factors associated with the health event being investigated
    b. if populations are dissimilar, variable-specific rates should be computed and compared
    c. rates can be adjusted for effect of a confounding variable in order to obtain undistorted view of effect that other variables have on risk
    d. adjustment of rates when comparing populations is called standardization and yields standardized or adjusted rates
    e. techniques of standardization
        1) direct
        2) indirect

3. direct standardization
    a. obtained by averaging specific rates for population, using distribution of a selected standard population as averaging weights
    b. adjusted rate represents what the crude rate would have been in study population if that population had same distribution as standard population with respect to variable for which adjustment or standardization was carried out
    c. specific rates are used directly in the calculation

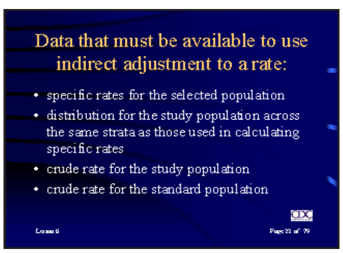
Return to Lesson 6

d. data which must be available to use direct adjustment
   1) specific rates for study population
   2) distribution for the selected standard population across the same strata as those used in determining the specific rates

4. indirect standardization
   a. calculated for a study population by averaging the specific rates for a select standard population, using distribution of the study population as weights
   b. when should one use indirect standardization



      1) when any of specific rates in study population are unavailable
      2) when such small numbers exist in categories of strata that data are unreliable
   c. indirect standardization is used more frequently than direct standardization
   d. data which must be available to use indirect adjustment to a rate



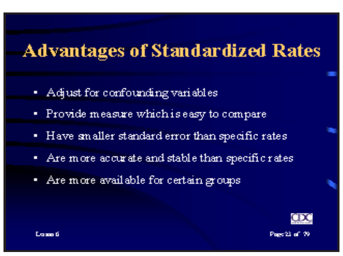
      1) specific rates for the selected standard population
      2) distribution for the study population across the same strata as those used in calculating specific rates
      3) crude rate for study population
      4) crude rate for standard population
   e. standardized mortality rate (SMR)
      1) special application of indirect standardized rate
      2) used when health event of interest is death

Return to Lesson 6

3) it is the number of deaths occurring in a study population or subpopulation
4) is expressed as a percentage of the number of deaths expected to occur if the given population and the selected standard population had the same specific rates
5) it is an indirect, age-adjusted ratio calculated as the indirect standardized mortality rate for the study population, divided by the crude mortality rate for the standard population

5. choice of standard population
   a. factors for consideration when choosing a standard population
      1) select a population that is representative of the study populations being compared
      2) understand how choice of a standard population affects directly standardized rates
   b. choices for a standard population
      1) combined or pooled population of overall population to be studied
      2) a large population that included the two study groups
      3) a hypothetical population that is representative of the two study groups

6. to standardize or not to standardize

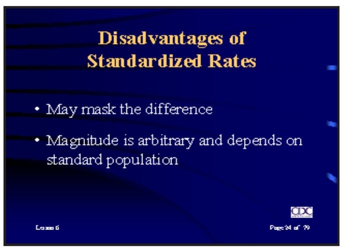


   a. reasons to present standardized rates
      1) standardization adjusts for confounding variables to yield a more realistic view of the effect of other variables on risk

Return to Lesson 6

2) a summary measure for a population is easier to compare with similar summary measures than are sets of specific rates
3) a standardized rate has a smaller standard error than any of the specific rates
4) specific rates may be imprecise or unstable because of sparse data in the strata
5) specific rates may be unavailable for certain groups of interest

b. disadvantages of standardization

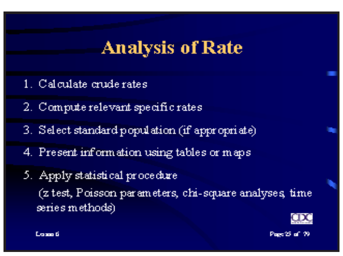


1) standardized rate can mask the difference and no single summary measure will reveal differences
2) magnitude is arbitrary and depends entirely on standard population

c. a compromise to the use of a summary measure versus a set of specific measures is to use specific rates but to eliminate or combine categories to minimize the number of rates required for comparison

7. rate standardization: practical exercise
   a. example demonstrates how standardized rates can be misleading if they are not properly scrutinized
   b. exercise - (see pages 108-110 in textbook)

D. Analysis of rates



1. begin with calculation of crude rates

2. proceed to subsequent computation of relevant specific rates

3. if appropriate, a standard population can be chosen to determine standardized rates

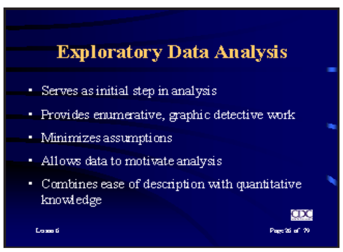
Return to Lesson 6

4. tables and maps are important ways to present rates at different times or locations

5. statistical procedures available to analyze data
    a. z test can be used for inference on a single proportion
    b. z or chi squared test can be used to assess the difference between two proportions
    c. Poisson parameters can be used to compare two rates (small numbers)
    d. time-series methods (smoothing, Box-Jenkins, Kalman filter approaches) can be used to analyze rates

IV. Exploratory data analysis (EDA)

A. EDA



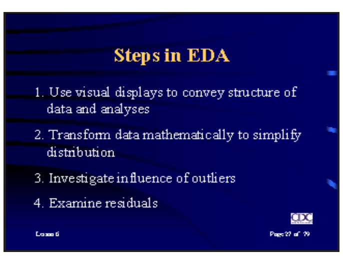
1. the initial step in any analysis

2. is enumerative, numeric, graphic detective work

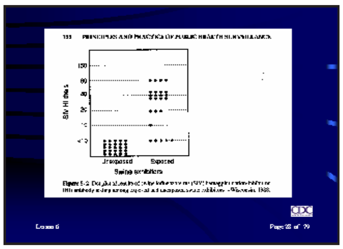
3. is the application of a set of techniques to a body of data to make data more understandable

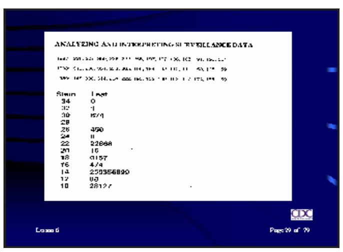
4. is a philosophy which minimizes assumptions, allows the data to motivate analysis, and combines ease of description with quantitative knowledge

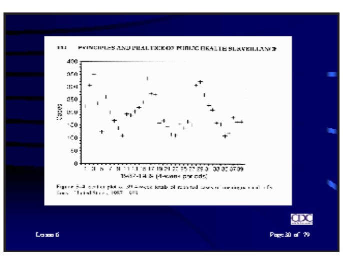
5. steps in the practice of EDA



Return to Lesson 6







d. examine the residuals (difference between the observed data and a fitted model) to provide additional insight into the data

6. can contribute to more timely detection of unusual observations, which may facilitate a quicker public health response

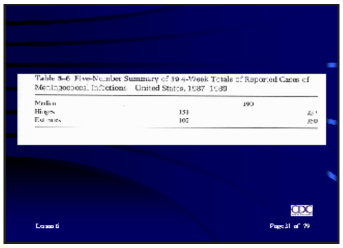
B. Techniques for data displays

1. dot plots
   a. a one-dimensional plot of the individual values of a set of numbers
   b. x-axis represents one or more categories of a noncontinuous variable
   c. y-axis represents range of values displayed by the observations
   d. observations with identical values are plotted side by side on the same horizontal plane

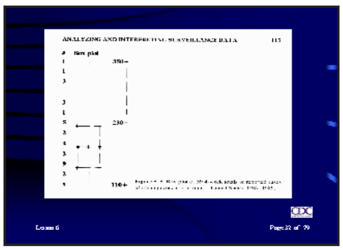
2. stem-and-leaf display
   a. graphic that allows the digits of the observation values to sort the numbers into numerical order for display
   b. is a variation of the conventional histogram
   c. each data value is split between a suitable pair of adjacent digits to form a set of leading digits and a set of trailing digits
   d. set of leading digits forms the stems, and the set of the first trailing digit from the data forms the leaves
3. scatter plot
   a. reveals relationship between two variables
   b. each observation comprises a pair of values
   c. observation is plotted by measuring the value of one variable on the horizontal axis and the value of the other on the vertical axis

Return to Lesson 6

C. Types of data summaries



1. five-number summaries
    a. simple display involving the median, hinge, and extreme values
    b. median is a measure of central tendency of data that splits an ordered data set in half
    c. hinges are a measure of the variability of the data and are the values in the middle of each half and are similar to quartiles
    d. extremes reflect the variability of the data and are the smallest and largest values in the data



2. box plots
    a. graphic representation
    b. 5-number summary with two ends of box representing the hinges and the line through the box representing the median
    c. can see quickly the median level, variability, symmetry of data

D. Transformations

1. powerful tool that facilitates understanding of data implications

2. reasons to transform raw data
    a. to achieve symmetry
    b. to produce a straight-line relationship
    c. to allow use of an additive model
    d. to reduce variability
    e. to attain normally distributed data

3. not all data sets can be transformed
    a. ratio of largest to smallest value in original data set is a simple indicator of whether a group of numbers will be affected substantially by transforming

Return to Lesson 6

b. if ratio is near 1, a transformation will not severely alter appearance of data
c. the further the ratio is from 1, the greater the need is for transformation to display and understand the data most simply
d. transformations are accomplished by raising each value of the data set to some power "p"

Return to Lesson 6

E. Smoothing

1. EDA techniques that summarize consecutive, overlapping segments of a series of data to produce a smoother curve

2. goal is to represent patterns in the data more clearly without becoming encumbered with detailed peaks and valleys

3. irregular components are smoothed so that the overall trend can be determined more readily

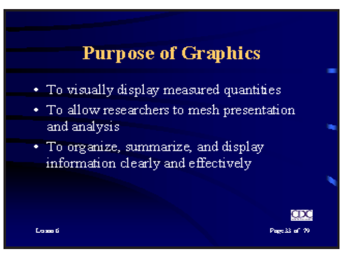
4. smoothing allows investigators to search for patterns that may otherwise be masked

5. smoothing is used on data series to explore relationship between 2 variables

6. simplest example is a moving average of three intervals

V. Data Graphics

A. Purpose of graphics



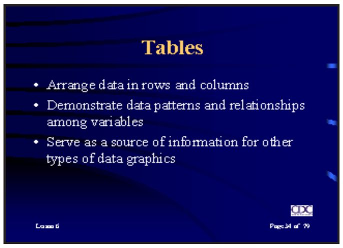
1. to visually display measured quantities

2. to allow researchers to mesh presentation and analysis

3. to organize, summarize, and display information clearly and effectively

4. selection of visual tools depends on data to be presented and purpose of presentation

Return to Lesson 6



B. Tables

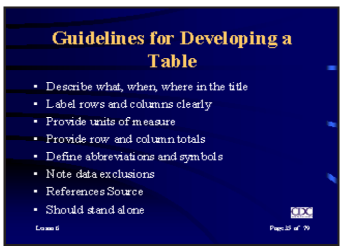
1. table arranges data in rows and columns

2. table is used to demonstrate data patterns and relationships among variables

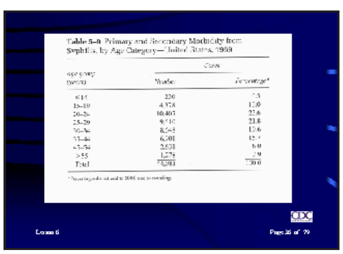
3. table serves as a source of information for other types of data graphics

4. table entries can be counts, means, rates, or other analytic measures

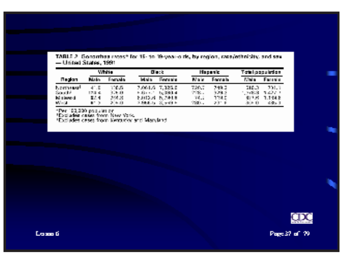
5. table should be simple and self-explanatory



6. guidelines in developing a table
   a. describe what, when, and where in a clear, concise table title
   b. label each row and column clearly
   c. provide units of measure for the data
   d. provide row and column totals
   e. define abbreviations and symbols
   f. note data exclusions
   g. if the data are not original, reference the source
   h. should be able to stand alone



7. one-variable tables
   a. is a frequency distribution by category for a single variable
   b. one of most basic tables

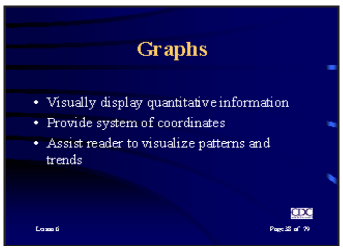


8. multivariable tables
   a. used when data are available on more than one variable
   b. also called contingency tables when all primary table entries are classified by each of the variables in the table

Return to Lesson 6

c. two-by-two contingency table
    1) used when two variables, each with two categories, are studied
    2) most frequently used type of table in epidemiologic analysis
    3) well-suited for analyzing case-control and cohort studies for which the categories of the variables are *case* and *control* (or *ill* and *well*) and *exposed* and *unexposed*

C. Graphs

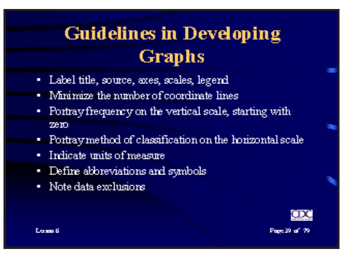


1. visual display of quantitative information

2. system of coordinates

3. primary analytic tools used to assist the reader to visualize patterns, trends, aberrations, similarities, and differences in data

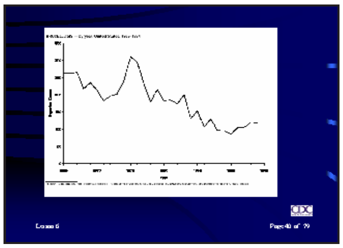
4. two-dimensional graphs are generally depicted along an x-axis (horizontal orientation) and y-axis (vertical orientation) coordinate system



5. guidelines in developing graphs
    a. keep it simple
    b. clearly and concisely label title, source, axes, scales, and legends to make graph self-explanatory
    c. clearly differentiate variables by legends or keys
    d. minimize the number of coordinate lines
    e. portray frequency on the vertical scale, starting at zero, and the method of classification on the horizontal scale

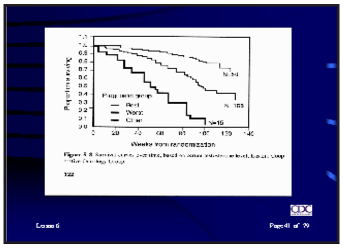
Return to Lesson 6

f. assure that scales for each axis are appropriate for data
g. clearly indicate scale division, any scale breaks, units of measure
h. define abbreviations and symbols
i. note data exclusions
j. if data are not original, reference the source

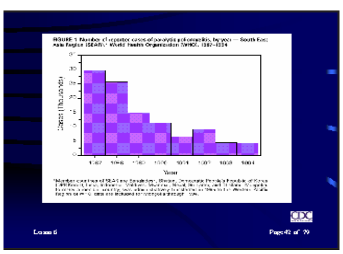


6. arithmetic-scale line graphs
   a. a graph in which equal distances along the *x* and *y* axes represent equal quantities along that axis
   b. typically used to demonstrate an overall trend over time rather than to focus on particular observation values
   c. helpful for examining long series of data or for comparing several data sets
   d. guidelines for developing arithmetic-scale line graphs
      1) length for the y-axis that is suitably proportional to that of the x-axis
      2) a common recommendation is a 5:3 x:y-axis ratio
      3) identify maximum y-axis value and round the value up slightly
      4) select an interval size that provides enough detail for the purpose of the graph
      5) scale breaks can be used for either or both axes if range of data is excessive but avoid misrepresentation or misinterpretation of data

Return to Lesson 6



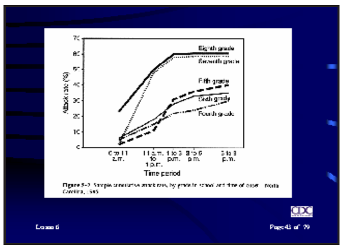
7. semilogarithmic-scale line graphs
    a. one axis is measured on an arithmetic scale (usually the x-axis) and the other is measured on a logarithmic scale
        1) logarithm is the exponent expressing the power to which a base number is raised
        2) example: log 100-log 10(squared)-2 for base 10
        3) axis portraying the logarithmic scale on semilog graph paper is divided into several cycles, with each cycle representing an order of magnitude and values 10 times greater than the preceding cycle
    b. semilogarithmic-scale line graphic is particularly valuable when examining the rate of change in surveillance data
    c. useful when large differences in magnitude or outliers occur because this type of graph allows the plotting of wide ranges of values
    d. slope of line indicates the rate of change



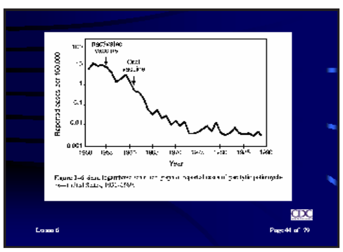
8. histograms
    a. a graph in which a frequency distribution is represented by adjoining vertical bars
        1) area represented by each bar is proportional to the frequency for that interval
        2) scale breaks should never be used because they misrepresent data
    b. constructed with equal- and unequal-class intervals; however, histograms for the latter category of data are difficult to construct

Return to Lesson 6

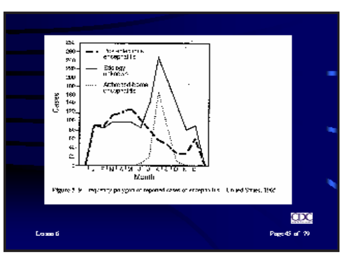
c. epidemic curve
    1) special type of histogram
    2) time is the variable plotted on the x-axis
    3) represents the occurrence of cases of a health problem by date of onset during an epidemic



9. cumulative frequency curves
    a. used for continuous and categorical data
    b. plots the cumulative frequency on the y-axis and the value of the variable on the x-axis
    c. cumulative frequencies can be expressed either as the number of cases or as a percentage of total cases
    d. for categorical data, the cumulative frequency is plotted at the right-most end of each class interval
    e. for percentages, cumulative frequency curve allows easy identification of medians, quartiles, and other percentiles of interest

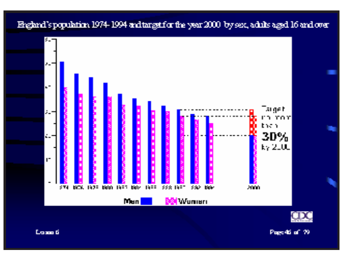


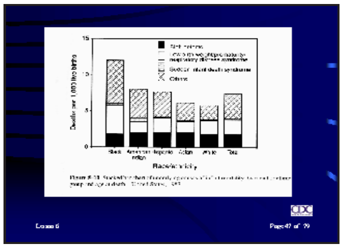
10. survival curve
    a. useful in a follow-up study for graphing the percentage of subjects remaining until an event occurs in the study
    b. x-axis represents time
    c. y-axis is percentage surviving



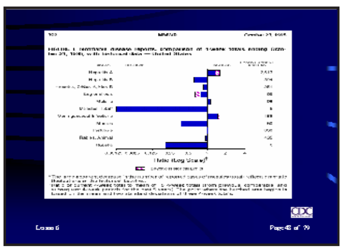
11. frequency polygons
    a. constructed from a histogram
    b. connect the midpoints of the class intervals with a straight line
    c. useful for comparing frequency distributions from different data sets

12. charts
    a. useful for illustrating statistical information
    b. useful for comparing magnitudes of events in categories of a variable
    c. bar charts
        1) simple and effective way to present comparative data
        2) uses bars of the same width to represent different categories of a factor
        3) comparison of categories is based on linear values since the length of a bar is proportional to the frequency of the event in that category
        4) scale breaks could cause the data to be misinterpreted and should not be used
        5) bars from different categories are separated
        6) can be depicted vertically or horizontally
        7) usually arranged in ascending or descending length
        8) variations of bar chart include grouped, stacked, and deviation bar charts
        9) grouped bar chart
            a) also called multiple-unit bar chart
            b) compares units within categories
            c) generally number of units within a category is limited to 3 for effective presentation and understanding

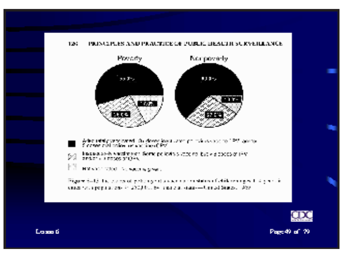




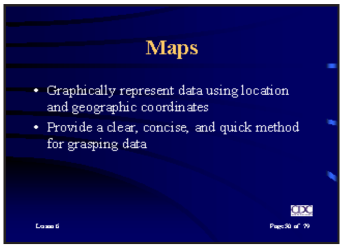
10) stacked bar chart
   a) also used to compare different groups within each category of a variable
   b) different groups are differentiated not with separate bars, but with different segments within a single bar for each category
   c) distinct segments are illustrated by different types of shading, hatching, or coloring, which are defined in a legend

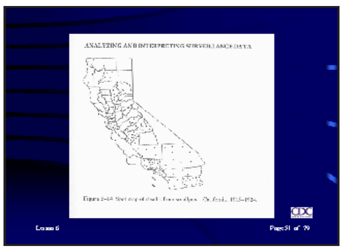


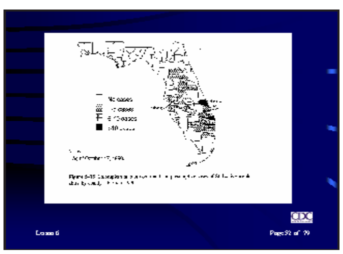
11) deviation bar chart
   a) illustrates differences in either direction from a baseline
   b) confidence interval-like portion can be added



12) pie chart
   a) represents the different percentages of categories of a variable by proportionally sized pieces of pie
   b) pieces are usually denoted with different colors or shading
   c) percentages are written inside or outside the pieces to allow the reader to make accurate comparisons







D. Maps

1. graphic representations of data using location and geographic coordinates

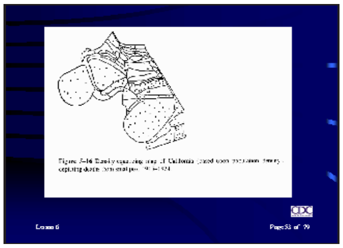
2. provides a clear, quick method for grasping data

3. effective for readers who are familiar with the physical area being portrayed

4. spot maps
   a. produced by placing a dot or other symbol on the map where health condition occurred or exists
   b. different symbols are used for multiple events at a single location
   c. spot map does not provide a measure of risk since population size is not taken into account
   d. spot map is beneficial for displaying geographic distribution of an event

5. chloropleth maps
   a. also called shaded or area map
   b. statistical map involving different types of shading, hatching, or coloring to portray range-graded values
   c. useful for depicting rates of a health condition in specific areas
   d. each area is shaded uniformly regardless of any demographic differences within an area
   e. can give false impression of abrupt change in number or rate of a condition across area boundaries when, in fact, a gradual change may have occurred from one area to the next

Return to Lesson 6



6. density-equalizing maps
    a. also called a rubber map
    b. transforms actual geographic coordinates to produce an artificial figure in which area or population density is equal throughout the map
    c. density-equalizing maps correct for the confounding effect of population density
    d. are useful in analyzing geographic clusters of public health events
    e. several algorithms exist to transform coordinates of maps
    f. transformation routine
        1) should define a continuous transformation over the map domain
        2) should solve for the unique solution that minimizes map distortion
        3) should accept optional constraints
        4) should avoid overlapping of transformed areas

VI. Interpretation of surveillance data

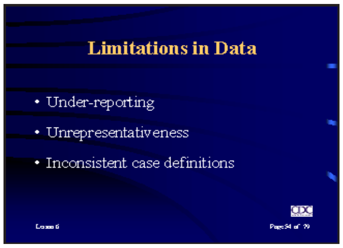
A. Factors in interpreting data

1. data need to be interpreted in the context of our understanding of the etiology, epidemiology, and natural history of the disease and injury

2. art of conducting surveillance lies in interpreting what the data say

3. key to interpretation lies in knowing the limitations of the data

4. correlation does not equal causation



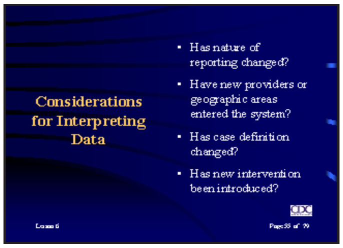
B. Limitations in data

1. under-reporting
    a. is inevitable because most surveillance systems are based on conditions reported by health care providers
    b. disease trends however, by time, place, and person frequently are detected even with incomplete data
    c. the rarer the disease, the more important completeness of reporting becomes

2. unrepresentativeness of reported cases
    a. health conditions are not reported randomly
    b. reporting biases can distort interpretation
    c. when possible, adjusting for skewed reporting will allow investigators to obtain a more accurate picture of the occurrence of a health problem
    d. collecting data from multiple sources may provide ways to improve the representativeness of the information

3. inconsistent case definitions
    a. different practitioners frequently use different case definitions for health problems
    b. the more complex the diagnostic syndrome, the greater the difficulty in reaching consensus on a case definition
    c. case definitions are frequently adjusted as understanding of newly emerging problems occur

Return to Lesson 6



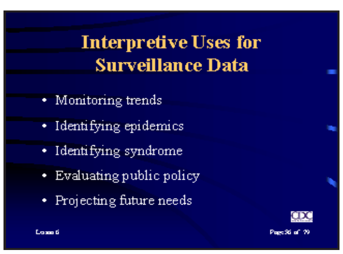
C. Approach to interpretation

1. factors to consider concerning variations in surveillance data
   a. has the nature of reporting changed?
   b. have providers or new geographic areas entered the surveillance system?
   c. has the case definition changed?
   d. has a new intervention, such as screening or therapy, been introduced?

2. consistency among different surveillance systems
   a. most crucial factor affecting surveillance
   b. if different surveillance systems show similar trends, greater likelihood that the effect is real
   c. select surveillance system which represents the highest quality local information

3. facilitation of interpretation
   a. design formats to determine whether the number of reported cases of a health problem for a specified reporting period differs from that of a previous period
   b. CDC's *Morbidity and Mortality Weekly Report (MMWR)* is a good example of a "user-friendly" format

VII. Interpretative uses for surveillance data



A. Monitoring trends

1. baseline level of the health condition being monitored reflects any variation in its occurrence over time

2. useful in infectious or noninfectious diseases

   a. progressive decline of tuberculosis in 20th century (until recently) and increase in AIDS throughout the 1980s reflect this monitoring function
   b. example: is relevant to assessing events associated with reproductive health
   c. example: projecting the decline in cirrhosis-related mortality in the presence of lower levels of alcohol use
   d. example: projecting decreased rates of mortality from cervical cancer associated with an increase in the prevalence of hysterectomy

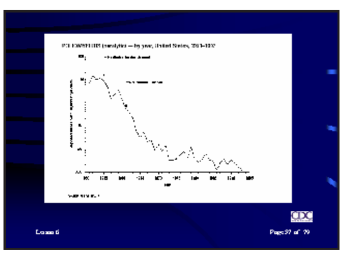
B. Identifying epidemics

1. most effective in circumstances when the health condition occurs over a wide geographic area or gradually over time

2. time-place-person links among cases probably would not be recognized by individual practitioners

3. example: when laboratory monitoring of unusual serotype of antibiotic-resistance patterns identify outbreaks of specific microorganisms that might otherwise have gone unnoticed

4. a surveillance system can function as an early warning signal for public health officials
   a. example: increases in numbers of cases of hepatitis B among military recruits provided the stimulus to intervene with drug prevention programs
   b. example: CDC's Birth Defects Monitoring System identified increases in renal agenesis during the 1970s and 1980s, which prompted an investigation



c. example: monitoring of regional trends in rubella and congenital rubella identified outbreaks among the Amish in 1989-1990

d. example: a national registry of antiabortion-associated violence documented an "epidemic" of attacks in the mid-1980s, which have varied depending on the level of prosecution allowed

C. Identifying new syndromes

1. Legionnaire's disease

2. AIDS

3. national scope of the epidemic of eosinophilia myalgia syndrome (EMS)

D. Evaluating public policy

1. surveillance data can assess the health impact of specific interventions or of public policy
   a. example: rapid fall in numbers of cases of poliomyelitis and measles after national vaccination campaigns
   b. example: adoption of a gun-licensing law coincided with an abrupt decline in firearm-related homicides and suicides in Washington, D.C.
   c. example: after legal abortion became widely available, deaths from illegal abortion decreased markedly

2. evaluation becomes increasingly suspect when several factors contribute to the occurrence of the disease or health condition being monitored

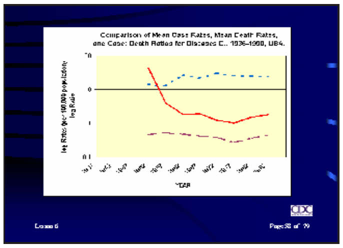
3. if only a portion of population accepts an intervention, analysis and interpretation of surveillance data are made more difficult

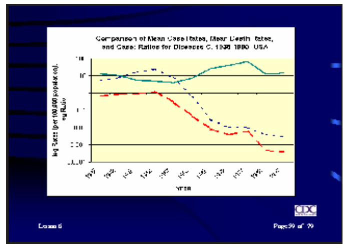
E. Projecting future needs

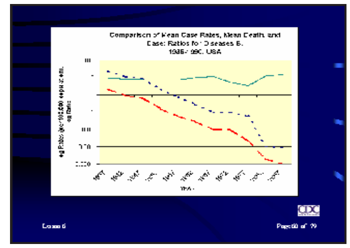
1. mathematical models based on surveillance data can be used to project future trends
   a. helps health officials determine the eventual need for preventive and curative services
   b. such modeling assisted in estimating the impact of AIDS on the U.S. health-care system in the 1990s
   c. example: projections addressed the demand for AZT and for the requirements for hospital care for those with HIV-related disease

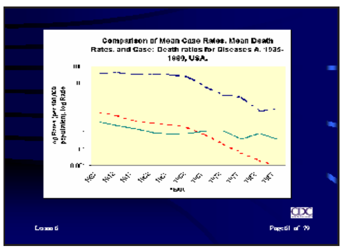
2. models based on surveillance data can predict the decline of morbidity and mortality when there are changes in risk factors among the population at risk
   a. example: projecting the decline in cardiovascular disease on the basis of decreased smoking of cigarettes

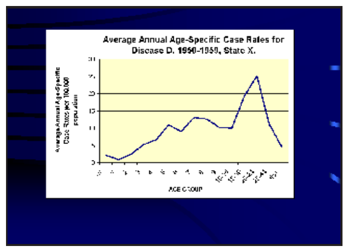
Return to Lesson 6

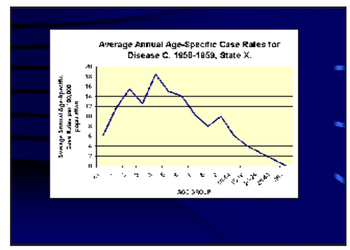

Comparison of Mean Case Rates, Mean Death Rates, and Case: Death Ratios for Diseases C., 1936-1990, USA.
YEAR
Lesson 6
Page 58 of 79

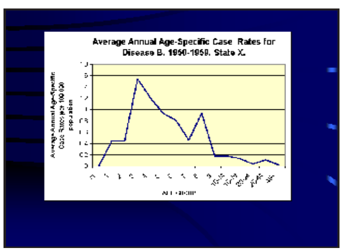

Comparison of Mean Case Rates, Mean Death Rates, and Case: Ratios for Diseases C. 1936-1990 USA
YEAR
Lesson 6
Page 59 of 79

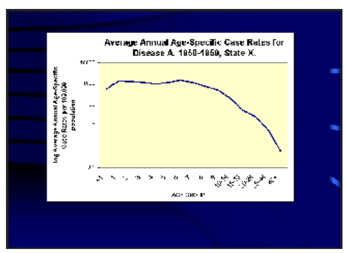

Comparison of Mean Case Rates, Mean Death, and Case: Ratios for Diseases B. 1936-1990. USA
Lesson 6
Page 60 of 79

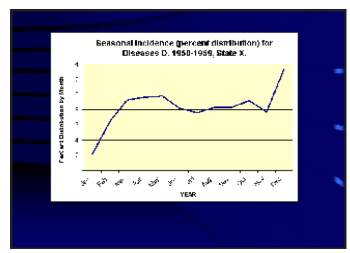

Comparison of Mean Case Rates, Mean Death Rates, and Case: Death ratios for Diseases A. 1936-1990, USA.
Lesson 6
Page 61 of 79

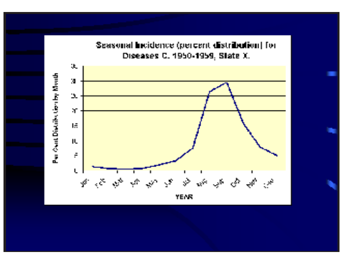

Average Annual Age-Specific Case Rates for Disease D. 1950-1959, State X.
AGE GROUP

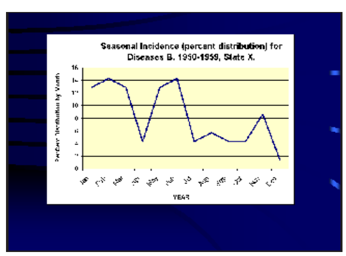

Average Annual Age-Specific Case Rates for Disease C. 1950-1959, State X.
AGE GROUP

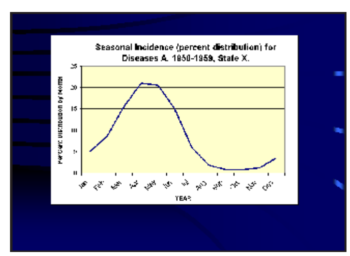

Average Annual Age-Specific Case Rates for Disease B. 1950-1959, State X.

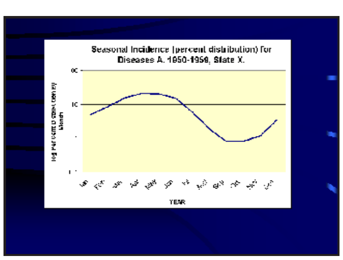

Average Annual Age-Specific Case Rates for Disease A. 1950-1959, State X.

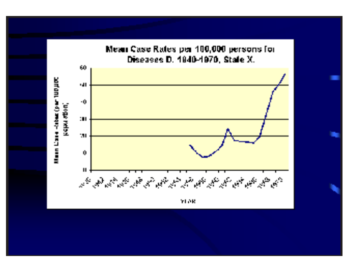

Seasonal Incidence (percent distribution) for Diseases D. 1950-1959, State X.
YEAR

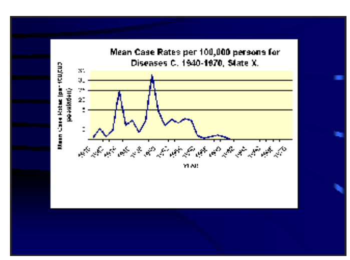

Seasonal Incidence (percent distribution) for Diseases C. 1950-1959, State X.
YEAR

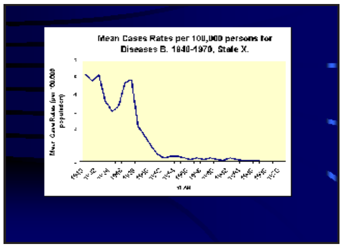

Seasonal Incidence (percent distribution) for Diseases B. 1950-1959, State X.
YEAR

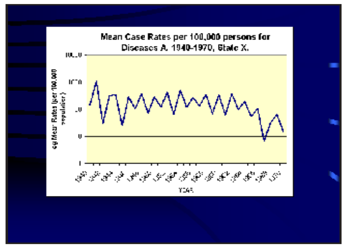

Seasonal Incidence (percent distribution) for Diseases A. 1950-1959, State X.
YEAR

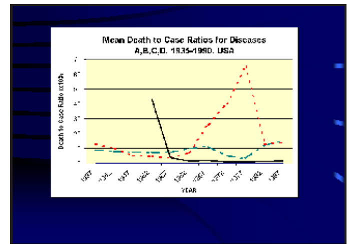

Seasonal Incidence (percent distribution) for Diseases A. 1950-1959, State X.
YEAR

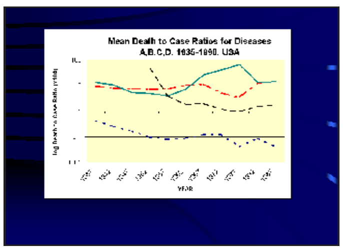

Mean Case Rates per 100,000 persons for Diseases D. 1940-1970, State X.

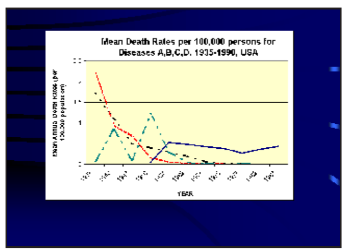

Mean Case Rates per 100,000 persons for Diseases C. 1940-1970, State X.

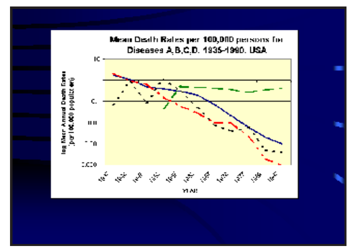

Mean Cases Rates per 100,000 persons for Diseases B. 1940-1970, State X.
YEAR

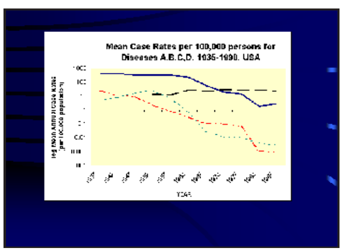

Mean Case Rates per 100,000 persons for Diseases A. 1940-1970, State X.
YEAR


Mean Death to Case Ratios for Diseases A,B,C,D. 1935-1990. USA
YEAR


Mean Death to Case Ratios for Diseases A,B,C,D. 1935-1990. USA
YEAR


Mean Death Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990, USA
YEAR


Mean Death Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990. USA
YEAR


Mean Case Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990. USA
YEAR